Diesel Fuel Effects on Emissions: Analysis Approach

EPA Workshop

Ann Arbor, MI

August 28, 2001

Robert L. Mason

Janet P. Buckingham

Southwest Research Institute

Overview

- Introduction
- Initial Modeling Approach
- Mixed Model Methodology
- Unified Model Approach
- Model Performance

Overall Approach

- Construct database from existing reports and studies
  - 1777 observations on 73 engines, 300 fuels, and 16 engine tech groups
- Model diesel emissions as a function of both engine and fuel properties

Initial Considerations

- Pre-standardize fuel properties
  - Facilitated coefficient comparisons
  - Reduced potential correlations between linear and squared fuel terms
- Transform emissions
  - Chose natural log transformation
  - Reduced variation & improved fit

Repeat Measurements

- Some studies had multiple repeat tests
- Some had single observations and no repeats
- Some had only averages of repeat tests
  - When the number of repeats in the average was unknown, the data for the average was repeated two times

Use of Repeats

- Initially considered limiting the number of repeats in the database
  - Constructed a file with no more than 4 repeats per engine-fuel combination
  - Randomly selected the 4 repeats
- Did this so as not to overweight data in the models

Initial Modeling Approach

- Fit individual tech group data
- Eigenvector analysis
  - Biased estimation procedure
  - Advantageous with strong collinearities
- Stepwise mixed model analysis
  - Candidate terms included 9 linear and 3 squared fuel properties

Mixed Model Methodology

- Contains both fixed and random effects
  - Thus labeled mixed model procedure
- Fuel properties are fixed effects
  - Controlled selection process for property values
- Engines are considered random effects
  - Engines are a sample from a population of possible engines

Form of Mixed Model

$$Y = X\beta + Zu + e$$

where

Y = emission

$X\beta$ = fixed fuel effects

Zu = random engine effects

e = random error terms

Assume u and e are normal with mean=0

Applicability of Mixed Model

- Provides predictor of aggregate emissions from overall population of engines
- Adds estimation of engine variance as well as error variance to model
- Accommodates unequal variances
- Allows nesting of fuel effects within each engine

Initial Results

- Eigenvector analysis deleted no fuel terms
  - Partitioning formula was not accurate
- Mixed model results for largest tech groups were somewhat similar
- Limited data for several tech groups
  - Some only contained a single engine
  - Had to choose terms to include in others

Unified Model Approach

- Two-step procedure was followed
- In step 1, stepwise regression was applied
- In step 2, a mixed model with a backwards-elimination was applied
- In both cases, hierarchical models were of interest

Use of Repeats

- Two data files were constructed
- First file contained average-repeat data
  - All repeat data were averaged
  - Singleton points were left alone
- Second file contained combined data
  - All data were included without restrictions and repeats were not averaged

Repeat Data Usage

- Average-repeat data file was used in the stepwise regression runs
  - Maintained equal weighting of the data points
- Combined-data file was used in mixed model runs
  - Unequal weighting not an issue since repeats aid in variance estimation

Engine-Fuel Interaction Terms

- In mixed model runs, linear fuel terms were nested within each engine term
- Done to determine if each engine had different fuel effects
- These interactions helped improve estimates of engine variation
  - Could affect significance of terms

Stepwise Regression Approach

- Fit model in a hierarchical manner after forcing engine terms in the model
- Sequentially considered fuel terms from following groups:
  - linear fuel, squared fuel
  - fuel-fuel interactions
  - techgrp-by-fuel interactions
  - techgrp-by-squared-fuel interactions

Stepwise Regression Procedure

- Analysis greatly simplified by automated stepwise process
- Provided quick identification of significant techgrp-by-fuel interactions
- Disadvantage included fact that some terms might be deleted early and not recognized later

Mixed Model Approach

- Built a hierarchical model using groups of candidate variables and backwards elimination within each group
- Began with terms in last stepwise model
- Engines and engine-fuel interactions were treated as random effects
- All other terms were treated as fixed effects

Mixed Model Backwards Procedure

Step1: Added tech group categorical variables corresponding to techgrp-by-fuel interactions in model

Step 2: Removed nonsignificant techgrp-by-fuel interactions in backward process

- Retained nonsignificant linear interaction if quadratic interaction was significant

Mixed Model Procedure

Step 3: Deleted nonsignificant tech group terms unless part of model hierarchy for techgrp-by-fuel interactions

Step 4: Deleted nonsignificant fuel-by-fuel interaction terms

Step 5: Deleted nonsignificant linear fuel terms, unless needed for model hierarchy

Mixed Model Results

Final model contained terms with significant coefficients, as well as terms with nonsignificant coefficients that were needed to maintain good model hierarchy

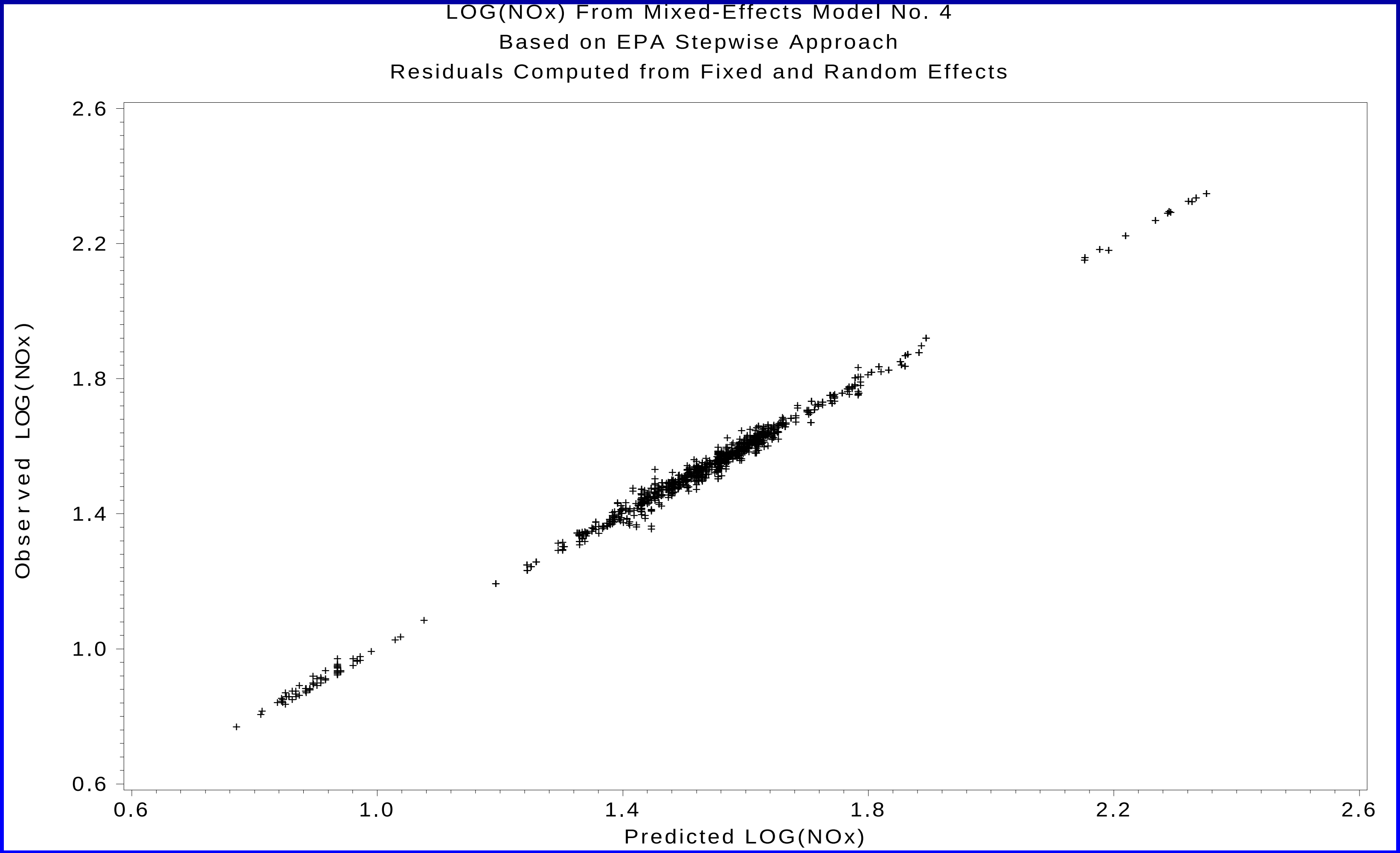
Model Performance

- Based on comparing Observed and Predicted % Change in Emissions relative to a national average base fuel

Obs % CE=100%(ObsFuel/ObsBase-1)

Prd %CE=100%(PrdFuel/PrdBase-1)

Model Performance Results



Model Performance Results

Cumulative % of the absolute differences between observed and predicted %CE that are between +/-10%

NO_x	99%
PM	81%
HC	47%

NOx Model Results

Variable	Stepwise Model	Mixed Model No Natural Cetane Interaction with Engines	Mixed Model
INTERCEPT	**1.61682**	**1.5326**	**1.5312**
NATURAL CETANE	**-0.00751 (p=0.0007)**	-0.00309 (p=0.0751)	-0.00033 (p=0.9047)
CETANE DIFFERENCE	**-0.01267**	**-0.01145**	**-0.01187**
TOTAL AROMATICS	**0.02779**	**0.02654**	**0.02679**
SPECIFIC GRAVITY	**0.01553**	**0.02195**	**0.02375**
SULFUR	0.00230	0.000932	0.000644
T10	-0.00101	0.004796	0.003553
T50	**-0.00978**	**-0.01396**	**-0.01459**

Summary

- Initial modeling approach led to use of combined database rather than individual tech group database
- Chosen methodology was combination of stepwise regression and mixed model
- Major advantage was the ability to predict aggregate emissions for overall population of engines represented by sample

Summary

- Prediction equations used to predict % change in emissions relative to a baseline fuel
- Model performance based on comparing observed and predicted % change in emissions